

Pengenalan Angka Terisolasi dengan Menggunakan Pemodelan HMM Melalui Ekstraksi Feature Mel-Cepstrum Filter Bank

Sri Jayadi (L2F302562)

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Diponegoro,
Jl. Prof. Sudharto, Tembalang, Semarang, Indonesia
E-mail: eeundip@indosat.net.id

Abstrak

Pengenalan suara adalah merupakan upaya agar manusia dan mesin dapat berkomunikasi dengan media suara. Pengenalan angka adalah merupakan salah satu bagian dari pengenalan suara. Salah satu contoh aplikasi dari pengenalan angka yaitu orang dapat menekan tombol telepon dengan menggunakan suara (*voice dial*).

Ketika suara digunakan sebagai media berkomunikasi antara manusia dan mesin, mesin tidak bisa begitu saja dapat mengenali suara yang diucapkan oleh manusia. Oleh karena itu suara perlu dimodelkan. Karena suara termasuk dalam suatu variabel yang acak maka digunakan suatu pemodelan statistik yaitu HMM (Hidden Markov Model). Dalam HMM suara dapat diasumsikan sebagai variabel acak yang dapat diperkirakan secara tepat. Sebelum suara dapat dimodelkan, suara harus melewati proses ekstraksi feature. Salah satu jenis ekstraksi feature adalah MFB (Mel-cepstrum filter bank).

Berdasarkan hasil pengujian dari sampel 1400 data yang diucapkan oleh 9 orang pria dan 5 orang wanita pengenalan angka terisolasi dengan menggunakan ekstraksi feature MFB dan pemodelan HMM didapatkan tingkat pengenalan tertinggi sebesar 97.64%, tingkat pengenalan ini dicapai pada state 11.

I. PENDAHULUAN

Suara merupakan salah satu bentuk alat komunikasi yang paling efektif. Selain efektif manusia juga sangat familiar dengan komunikasi menggunakan suara. Karena itu manusia berusaha mengembangkan suatu cara bagaimana agar manusia dan mesin dapat berinteraksi menggunakan suara. Supaya suara dapat digunakan sebagai media komunikasi antara manusia dan mesin suara perlu dimodelkan sehingga dapat dikenali.

Salah satu bentuk pemodelan suara adalah HMM (Hidden Markov Model). Dalam pemodelan ini suara dapat diasumsikan sebagai parameter acak yang dapat diperkirakan secara tepat, jadi HMM merupakan analisa probabilitas yang bersifat stokastik. Dengan HMM sinyal suara dapat dianalisis dan didapatkan nilai probabilitas yang terbaik sehingga bisa dikenali.

Tujuan dari Tugas Akhir ini adalah untuk dapat mengenali angka (0-9) dalam bentuk suara sehingga dihasilkan output berupa teks yang bersesuaian dengan input yang dimasukkan dan untuk menentukan jumlah state yang paling bagus dalam pengenalan suara

Pembatasan masalah pada Tugas Akhir pembuatan simulasi pengenalan angka dalam bentuk suara dengan HMM adalah:

- Analisis sinyal yang digunakan adalah Mel-Cepstrum Filter Bank.
- Data masukan merupakan angka dari 0 sampai 9 yang diucapkan dalam bahasa Indonesia oleh 9 orang pria dan 5 orang wanita yang direkam dengan frekuensi sampling 8 KHz, mono, 8 bit dan disimpan dalam berkas dengan ekstensi wav.
- HMM yang digunakan adalah Unimodal Continuous Hidden Markov Model (CHMM) dengan variasi state dari 2 sampai 20.
- Program simulasi ditulis dengan bahasa pemrograman Matlab 6.1.

II. KONSEP PENGENALAN ANGKA TERISOLASI DENGAN HIDDEN MARKOV MODEL

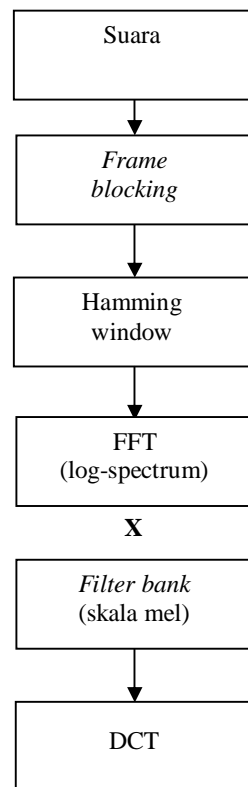
A. PENGENALAN SUARA

Pengenalan suara adalah upaya supaya suara dapat diidentifikasi sehingga dapat dimanfaatkan. Pengenalan suara secara umum dapat dibagi menjadi beberapa tahap yaitu: ekstraksi *feature*, pemodelan dan pengenalan.

Ekstraksi *feature* adalah upaya untuk memperoleh *feature* dari sinyal suara. Salah satu metode yang dapat digunakan untuk proses ekstraksi *feature* adalah Mel-Cepstrum. Setelah didapatkan *feature*, kemudian dilakukan pemodelan. Karena sinyal suara dapat dikarakteristikan sebagai variabel proses acak, sehingga untuk pemodelan ini dapat dilakukan dengan pemodelan statistik yaitu HMM (Hidden Markov Model). Dari pemodelan didapatkan parameter yang selanjutnya digunakan dalam proses pengenalan.

B. EKSTRAKSI FEATURE

Feature merupakan properti karakteristik dari sinyal suara. *Feature* dapat dibagi menjadi dua yaitu *feature* statik dan *feature* dinamik. *Feature* statik dapat diperoleh dengan analisa spektrum dengan metode mel-cepstrum, sedangkan untuk *feature* dinamik dapat diperoleh dari turunan koefisien mel-cepstrum. Diagram untuk mencari koefisien mel-cepstrum ditunjukkan pada Gambar 1.



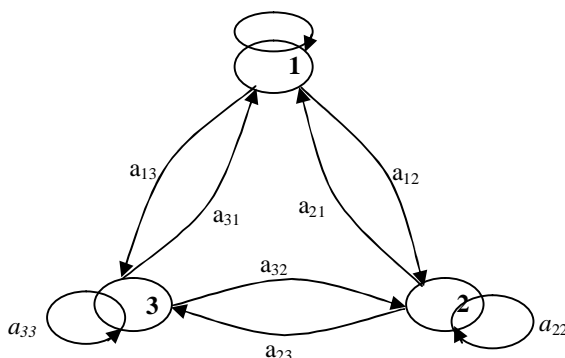
Gambar 1 Ekstraksi feature dengan mel-cepstrum

C. PEMODELAN DENGAN HIDDEN MARKOV MODEL (HMM)

HMM adalah analisa statistika yang memodelkan sinyal suara dan mencari bentuk kata yang paling sesuai. HMM berkembang pada akhir tahun 1960 dan awal tahun 1970. HMM berkembang dengan sangat cepat karena pemodelan ini sangat kaya dalam struktur matematika dan bisa digunakan untuk beragam aplikasi. HMM mengacu pada fungsi probabilitas rantai markov.

a. RANTAI MARKOV

Algoritma HMM didasari oleh model matematik yang dikenal dengan rantai markov. Rantai markov secara umum ditunjukkan pada Gambar 2.



Gambar 2 Rantai markov

Beberapa hal yang dapat dijelaskan tentang rantai markov yaitu:

- ▲ Transisi keadaan dari suatu keadaan tergantung pada keadaan sebelumnya.
 $P[q_t = j | q_{t-1} = i, q_{t-2} = k, \dots] = P[q_t = j | q_{t-1} = i]$
- ▲ Transisi keadaan bebas terhadap waktu.
 $a_{ij} = P[q_t = j | q_{t-1} = i]$

b. Definisi HMM

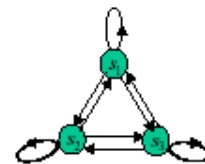
HMM adalah probabilitas teknik pencocokan pola yang observasinya dianggap sebagai output dari poses stokastik dan didasari rantai markov. HMM terdiri dari dua komponen: rantai markov keadaan terbatas dan output distribusi keadaan yang terbatas.

c. Tipe HMM

HMM dibagi menjadi dua tipe dasar yaitu HMM ergodic dan HMM Kiri-Kanan.

1. HMM *ergodic*.

Pada HMM model ergodic perpindahan keadaan satu ke keadaan yang lain semuanya memungkinkan, hal ini ditunjukkan pada Gambar 3.

Gambar 3 HMM model *ergodic*

2. HMM Kiri-kanan

Pada HMM kiri-kanan perpindahan keadaan hanya dapat berpindah dari kiri kekanan, perpindahan keadaan tidak dapat mundur ke belakang, hal ini ditunjukkan pada Gambar 4.



Gambar 4 HMM model kiri-kanan

Transisi keadaan untuk HMM kiri-kanan dapat dinyatakan dengan persamaan:

$$a_{ij} = 0, \quad j < i$$

d. Elemen HMM

Untuk HMM dengan observasi simbol diskret seperti model jambangan dan bola, elemen HMM adalah sebagai berikut:

1. N, Jumlah keadaan dalam model.
2. M, Jumlah simbol observasi tiap keadaan.
3. Distribusi keadaan transisi $A = \{a_{ij}\}$ dengan
 $a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N$
4. Distribusi probabilitas simbol observasi, $B = \{b_j(k)\}$ dengan

$$b_j(k) = P(o_t = v_k | q_t = j), \quad 1 \leq k \leq M$$

5. Distribusi keadaan awal

$$\pi_i = P[q_1 = i], \quad 1 \leq i \leq N$$

e. Prosedur Evaluasi

Salah satu permasalahan dasar dalam pemodelan HMM adalah menghitung probabilitas urutan observasi, $O=(o_1, o_2, \dots, o_T)$ yang diberikan oleh model λ .

1) Algoritma Forward

Jika variabel *forward* $\alpha_t(i)$, pada saat t dan keadaan i , maka didapat persamaan sebagai berikut

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = i | \lambda)$$

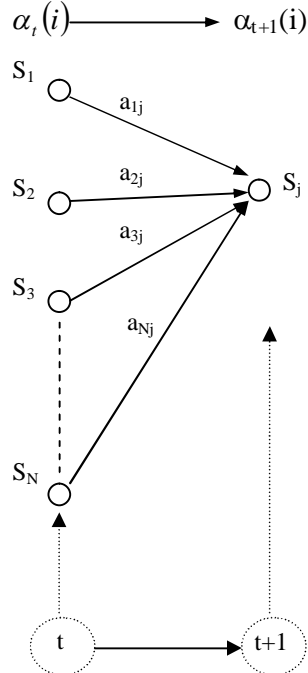
Fungsi probabilitas dapat diselesaikan untuk N keadaan dan observasi T secara iterasi:

▲ Inisialisasi

$$\alpha_t(i) = \pi_i b_i(o_1) \quad 1 \leq i \leq N$$

▲ Induksi

Proses induksi ditunjukkan pada Gambar 5.



Gambar 5 Fungsi probabilitas *forward*

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}),$$

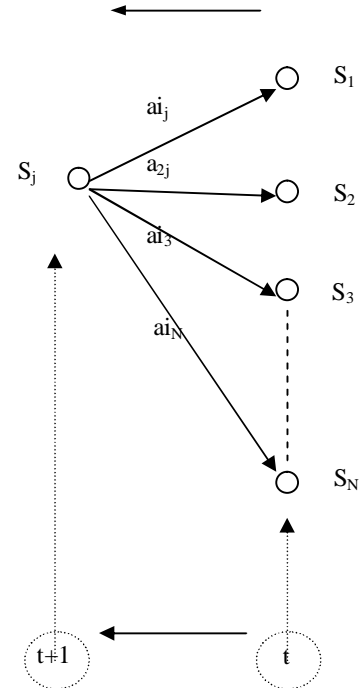
▲ Terminasi

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i)$$

2) Algoritma Backward

Algoritma ini mirip dengan algoritma *forward* tetapi keadaan mengalir kebelakang dari observasi terakhir saat T . Fungsi probabilitas *backward* $\beta_t(i)$ dapat didefinisikan sebagai berikut:

$$\beta_t(i) = P(o_{t+1} o_{t+2} \dots o_T | q_t = i, \lambda)$$



Gambar 6 Fungsi probabilitas *backward*

Dan dianalogikan dengan prosedur *forward* $\beta_t(i)$ dapat diselesaikan dengan dua langkah:

▲ Inisialisasi

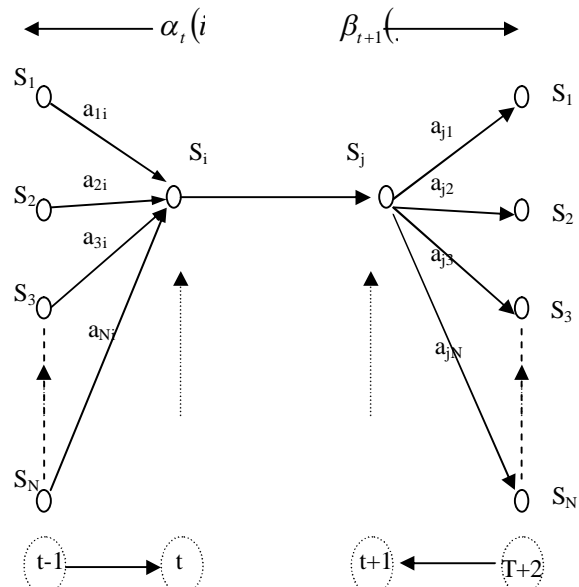
$$\beta_t(i) = 1 \quad 1 \leq i \leq N$$

▲ Induksi

$$\beta_t(i) = \sum_j a_{ij} b_j(o_{t+1}) \beta_{t+1}(j),$$

$$t = T-1, T-2, \dots, 1 \quad 1 \leq i \leq N$$

3) Algoritma Forward-backward.



Gambar 7 Fungsi probabilitas *forward-backward* untuk menentukan $P(O|\lambda)$

$$P(O | \lambda) = \sum_{i=1}^N \alpha_i(i) \beta_i(i)$$

4) Algoritma Viterbi

Algoritma viterbi hampir mirip dengan algoritma *forward*, perbedaannya hanya mengganti penjumlahan dengan pemaksimalan.

▲ Inisialisasi

$$\delta_i(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

▲ Rekursi

$$\delta_i(j) = \max_{1 \leq i \leq N} [\delta_{i-1}(i) a_{ij}] b_j(o_t),$$

$$2 \leq t \leq T$$

$$1 \leq j \leq N$$

▲ Terminasi

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

D. Prosedur Pelatihan.

Pelatihan berfungsi untuk mendapatkan model yang lebih baik. Algoritma yang digunakan untuk pelatihan adalah algoritma Baum-welch, yang disebut juga algoritma *forward-backward*.

$\hat{\pi}$ = Jumlah yang diharapkan saat keadaan awal adalah S_i

$$= \gamma_i(i)$$

Dengan demikian estimasi ulang parameter model dapat disimpulkan:

$$\hat{a}_{ij} = \frac{\text{Jumlah transisi yang diharapkan dari } S_i \text{ ke } S_j}{\text{Jumlah transisi yang diharapkan dari } S_i}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_t(j)}$$

$$\hat{b}_j(k) = \frac{\text{Jumlah } t \text{ yang diharapkan dalam } S_j \text{ dengan observasi } O_t = w_k}{\text{Jumlah yang diharapkan saat } S_j}$$

$$= \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} = \frac{\sum_{t: O_t = w_k} \alpha_t(j) \beta_t(j)}{\sum_{t=1}^T \alpha_t(j) \beta_t(j)}$$

Setelah estimasi ulang parameter model, maka akan didapatkan model lain $\hat{\lambda}$ yang lebih mirip dibandingkan λ , yang dihasilkan dari urutan observasi O . Hal ini berarti

$$P(O | \hat{\lambda}) > P(O | \lambda)$$

E. Penskalaan

Penskalaan sangat penting dalam implementasi HMM karena untuk menghindari hasil yang tak terhingga dalam penghitungan probabilitas. Teknik penskalaan, pertama adalah dengan mendefinisikan koefisien penskalaan $c(t)$:

$$c(t) = \frac{1}{\sum_{i=1}^N \alpha_i(i)}$$

Teknik penskalaan dapat digunakan untuk menentukan $\log P(O/\lambda)$ dengan metode sebagai persamaan:

$$C_T = \prod_{t=1}^T c_t = \frac{1}{P(O | \lambda)}$$

Kedua sisi dilogkan menjadi

$$\log \left(\prod_{t=1}^T c_t \right) = -\log [P(O | \lambda)]$$

dengan menggunakan sifat log didapat

$$\log [P(O | \lambda)] = -\sum_{t=1}^T \log(c_t)$$

F. Urutan Observasi yang Lebih dari Satu

Kelemahan dari HMM tipe kanan-kiri adalah urutan observasi tidak dapat dijadikan satu untuk pelatihan. Didefinisikan observasi dengan jumlah k sebagai berikut:

$$O = [O^1, O^2, O^3, \dots, O^k]$$

dengan

$$O^k = [O_1^k, O_2^k, O_3^k, \dots, O_{T_k}^k]$$

Tujuan pengaturan λ adalah untuk pemaksimalan $P(O|\lambda)$. Dalam kasus observasi lebih dari satu $P(O|\lambda)$ dapat didefinisikan dengan:

$$P(O | \lambda) = \prod_{k=1}^K P(O^k | \lambda)$$

lebih jelasnya

$$P(O | \lambda) = \prod_{k=1}^K P_k$$

G. Continuous Hidden Markov Model (CHMM)

Dalam CHMM parameter model juga berupa π , A , dan B , tetapi dalam bentuk yang berbeda. *Probability Density Function* (PDF) dari urutan observasi dianggap sebagai distribusi gaussian. Dianggap bahwa $b_i(O)$ adalah sesuai dengan rumus:

$$b_i(O) = \sum_{m=1}^M c_{im} n(O; \mu_{im}, U_{im}), \quad 1 \leq i \leq N$$

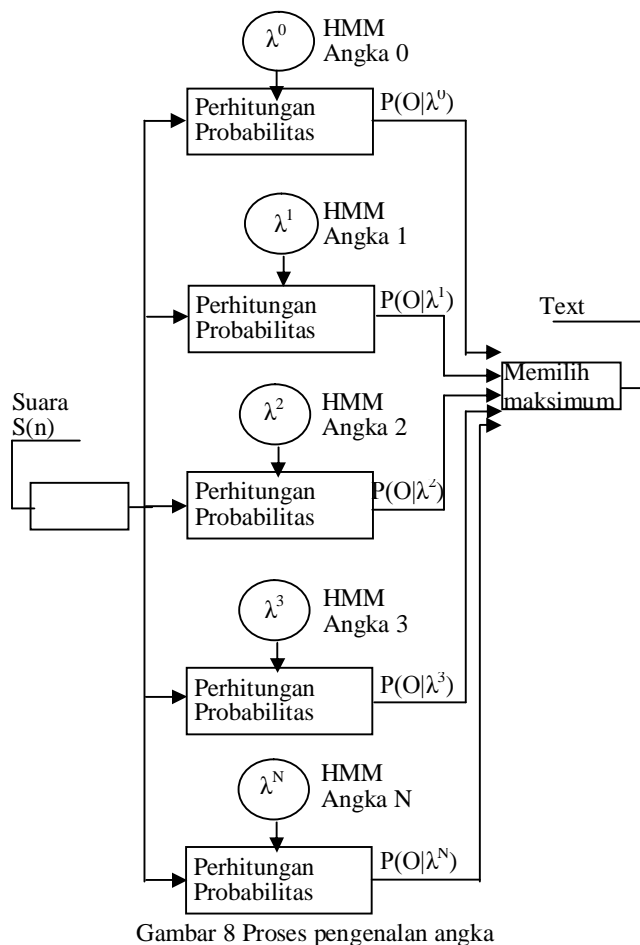
keterangan

c_{jm} = jumlah mixture pada keadaan i .

- n = distribusi pdf yang dianggap sebagai distribusi gaussian.
 μ_{im} = rata rata mixture ke m dalam keadaan i
 U_{im} = kovarian pada mixture ke m pada keadaan i
 O =urutan observasi dari *feature* vektor dengan dimensi d .
 M = jumlah mixture yang digunakan.
 N = jumlah keadaan.

H. PENGENALAN ANGKA TERISOLASI

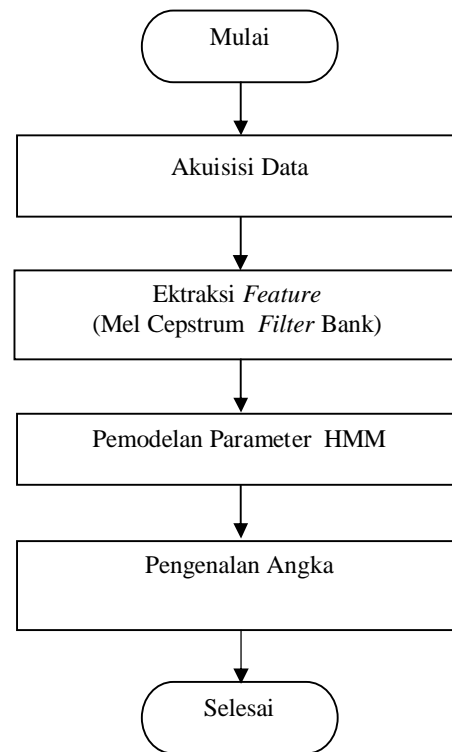
Yang dimaksud dengan terisolasi adalah masukan suara berupa angka tunggal. Pengenalan angka terisolasi ditunjukkan pada Gambar 8.



Sinyal suara $s(n)$ dimasukkan kedalam ekstraksi *feature* sehingga didapatkan vektor urutan observasi O . Kemudian dicari model HMM untuk tiap-tiap angka dan yang terakhir adalah menentukan kemungkinan terbesar urutan observasi terhadap λ , sehingga dari sini didapatkan keluaran yang bersesuaian dengan input.

III. PERANCANGAN DAN IMPLEMENTASI SISTEM

Secara umum pembuatan program simulasi ini mengikuti alur sesuai yang ditunjukkan pada Gambar 9.



Gambar 9 Alur pembuatan program simulasi pengenalan angka terisolasi

IV. ANALISA DAN PEMBAHASAN

Dari hasil pengujian untuk *state* 2 sampai 20 tingkat pengenalan terendah hampir semuanya terjadi pada angka 0 dan atau angka 9, karena ada data-data dari angka tersebut yang salah dikenali hal ini kemungkinan disebabkan karena angka-angka tersebut mempunyai nilai parameter yang hampir sama, tetapi mempunyai nilai probabilitas yang lebih rendah dibandingkan angka 6.

Apabila persentase pengenalan total dari setiap *state* ditabelkan maka akan didapatkan tabel seperti yang ditunjukkan pada Tabel 1.

Tabel 1 Persentase pengenalan semua *state*

State	Pengenalan (%)
2	89.43
3	91.29
4	94.57
5	95.71
6	96.00
7	96.50
8	96.86
9	97.14
10	96.50
11	97.64
12	97.43
13	97.00
14	97.57
15	96.71
16	97.57

Lanjutan Tabel 1.

State	Pengenalan (%)
17	97.00
18	97.57
19	97.00
20	97.14

Pengenalan terendah terjadi pada *state* 2 dan pengenalan tertinggi terjadi pada *state* 11. Pengenalan terendah terjadi pada *state* 2 karena dengan jumlah *state* yang sedikit cenderung tidak dapat membedakan setiap angka. Dari *state* 2 sampai 20 semuanya mempunyai kesalahan dalam pengenalan data hal ini disebabkan karena:

- Derau lingkungan sekitar.
Karena kondisi ruangan saat perekaman tidak kedap suara sehingga memungkinkan adanya suara-suara lain yang ikut terekam. Dengan adanya derau maka akan dihasilkan vektor *feature* yang tidak mencirikan sinyal asli sehingga menghasilkan nilai maksimum pada tempat yang salah. Untuk mengatasi hal ini diperlukan ruangan perekaman yang seminimal mungkin bebas derau.
- Derau sistem
Hal ini berkaitan dengan komputer yang digunakan sebagai media perekaman, saat belum ada suara window sound recorder sudah menunjukkan ada sinyal yang masuk, sinyal tersebut berarti derau. Untuk mengatasi hal ini dapat dilakukan dengan mengganti kartu suara dengan kualitas yang bagus.
- Karakteristik sinyal suara.
Walaupun angka diucapkan oleh orang yang sama, tetap mempunyai karakteristik yang berbeda-beda. Setiap kali angka yang diucapkan selalu saja mempunyai ciri yang berbeda, baik itu panjang-pendek, keras-pelan dan lain lain. Hal ini dapat diatasi dengan megusahakan pengucapan yang sama atau dengan memperbanyak data latihan. Dengan memperbanyak data latihan maka diharapkan didapatkan parameter yang lebih baik, sehingga tingkat pengenalan lebih tinggi.
- Letak mikropon.
Karena perekaman dilakukan secara bertahap sehingga dalam peletakan mikropon tidak sama. Jarak dan sudut mikropon sangat mempengaruhi pada sinyal suara yang dihasilkan. Untuk mengatasi hal ini dapat dilakukan dengan mengatur mikropon jarak dan sudut mikropon yang sebisa mungkin sama.

V. KESIMPULAN

Berdasarkan uraian pada bab-bab sebelumnya dapat ditarik kesimpulan sebagai berikut:

1. Tingkat pengenalan terendah dicapai untuk jumlah *state* 2 dengan prosentase pengenalan sebesar 89.43% dan tingkat pengenalan tertinggi dicapai untuk *state* 11 dengan prosentase pengenalan sebesar 97.64 %.
2. Kesalahan pengenalan terbesar untuk semua *state* terjadi karena adanya kesalahan dalam pengenalan angka 0 dan atau angka 9.

3. Dengan menggunakan ekstraksi *feature* mel-cepstrum filter bank dan pemodelan unimodal CHMM akan memberikan akurasi berkisar pada angka 96.14% untuk pengenalan angka terisolasi.

VI. SARAN

1. Sebaiknya diteliti lebih lanjut tentang pengenalan dengan kondisi lingkungan yang sebenarnya.
2. Untuk data latihan supaya diperbanyak, agar parameter yang didapat lebih baik sehingga tingkat pengenalan lebih akurat.
3. Estimasi ulang untuk parameter HMM sebaiknya dilakukan sampai nilai log likelihood mencapai konvergen, artinya nilai loglikelihood dengan nilai loglikelihood sebelumnya hampir tidak ada perubahan.
4. Untuk meningkatkan akurasi pengenalan perlu adanya suatu sistem yang dapat menekan derau seminimal mungkin.
5. Untuk pengenalan sebaiknya dikembangkan dengan menggunakan variasi-variasi HMM yang lain.

DAFTAR PUSTAKA

1. Abdulla, Waleed H., and Kasabov, Nikola K., "*The Concepts of Hidden Markov Model in Speech Recognition*", Knowledge Engineering Lab. Information Science Department University of Otago, New Zealand, 1999.
2. Cappe, Olivier., *H2M : A Set of Matlab/Octave Functions for The EM Estimation of Mixtures and Hidden Markov Models*, ENST dpt. TSI / LCTI (CNRS-URA 820), Paris, 2001.
3. Eidenbenz, Stephan, *Bioinformatics: Hidden Markov Model, Institute of Theoretical Computer Science*, Zurich.
4. Gold, Ben and Morgan, Nelson, *Speech Audio Signal Processing: Processing and Perception of Speech and Music*, John Wiley and Sons, USA, 2000.
5. Gu Liang and Liu, Ren Sheng, "*The Application of Optimmmization In Feature Extraction of Speech Recognition*", Department of Electronic Engineering Tsinghua University, Beijing.
6. Kinnunen, Tomi., Karkkainen, Ismo, and Franti, Pasi, "*Is Speech Data Clustered ? - Staistical Analysis of Cepstral Features*", University Joensuu, Department of Computer Science, Finland.
7. Loizou, Philip, "*Colea: A Matlab Software Tool for Speech Analysis*", Dept. of applied Science University of Arkansas at Little Rock, Little Rock, 1998.
8. L. R. Rabiner, "*A Tutorial on Hidden Markov Models and Selected Aplication in Speech Recognition*", vol. 77, no. 2, pp. 257-286, 1989.
9. Manolakis, Dimitris G., Ingle, Vinay K., and Kogon, stephen M., *Statistical and Adaptive Signal Processing*, McGraw-Hill, 2000.
10. Peebles, Peyton Z., Jr., *Probability, Random Variables, and Random Signal Principles*, McGraw-Hill, Inc., 2001.
11. Pekar, Darko, Obradoviae, Radovan, and Deliae, vlado, "*Connected Word Recognition*", School of

- Engineering, university of Novi Sad, Yugoslavia, 2001.
12. Proakis, John G., and Manolakis Dimitris G., **Pemrosesan Sinyal Digital; prinsip, algoritma dan aplikasi**, PT Prenhallindo, Jakarta, 1997.
 13. Rabiner, Lawrence and Juang, Biing-Hwang, **Fundamentals of Speech Recognition**, PTR Prentice Hall, 1993.
 14. Richardson, Matt, Bilmes, Jeff and Diorio, Chris, **"Hidden - Articulator Markov Models for Speech Recognition"**, University of Washington, Washington.
 15. Sarikaya, Ruhi, and Hansen, John H. L., **"Analysis of The Root-Cepstrum for Acoustic Modeling and Fast Decoding in Speech Recognition"**, Robust Speech Processing Laboratory Center for Spoken Language Research University of Colorado at Boulder, Boulder.
 16. Zhao, Yunxin, Wakita, Hisashi, and Zhuang Xinhua, **"An HMM Based Speaker-Independent Continuous Speech Recognition System with Experiment on the TIMIT DATABASE"**, Proc. ICASSP, Vol. I, pp 333-336, Toronto Canada, 1991.

Sri Jayadi,

Lahir di Pati 18 Oktober 1976. Saat ini sedang menyelesaikan Tugas Akhir untuk meraih gelar tingkat strata 1 (S1) di Jurusan Teknik Elektro Universitas Diponegoro Semarang dengan konsentrasi Telekomunikasi.

Menyetujui,

Dosen Pembimbing I,

Dosen Pembimbing II,

Wahyudi, ST, MT
NIP. 132 086 662

Achmad Hidayatno, ST, MT
NIP. 132 137 933